TryHackMe: Google Dorking

Google Dorking

Google hacking, also named Google Dorking, is a hacker technique that uses Google Search and other Google applications to find security holes in the configuration and computer code that websites are using. It is a technique that utilizes advanced search operators to uncover information on the internet that may not be readily available through standard search queries.

Offensive Security

Offensive Security involves breaking into computer systems, exploiting software bugs, and finding loopholes in applications to gain unauthorized access. The goal is to understand hacker tactics and enhance our system defenses. It aims to identify and exploit system vulnerabilities to enhance security measures. This includes exploiting software bugs, leveraging insecure setups, and taking advantage of unenforced access control policies, among other strategies. Red teams and penetration testers specialize in these offensive techniques.

`gobuster -u http://fakebank.thm -w wordlist.txt dir`

In the command above, -u is used to state the website we're scanning, -w takes a list of words to iterate through to find hidden pages.

If you were a penetration tester or security consultant, this is an exercise you'd perform for companies to test for vulnerabilities in their web applications and find hidden pages to investigate for vulnerabilities.

Here is a short description of a few offensive security roles:

- Penetration Tester - Responsible for testing technology products for finding exploitable security vulnerabilities.
- Red Teamer - Plays the role of an adversary, attacking an organization and providing feedback from an enemy's perspective.
- Security Engineer - Design, monitor, and maintain security controls, networks, and systems to help prevent cyberattacks.

Defensive Security

It is concerned with two main tasks:

- Preventing intrusions from occurring
- Detecting intrusions when they occur and responding properly

Some of the tasks that are related to defensive security include:

- User cyber security awareness: Training users about cyber security helps protect against attacks targeting their systems.
- Documenting and managing assets: We need to know the systems and devices we must manage and protect adequately.
- Updating and patching systems: Ensuring that computers, servers, and network devices are correctly updated and patched against any known vulnerability (weakness).
- Setting up preventative security devices: firewall and intrusion prevention systems (IPS) are critical components of preventative security. Firewalls control what network traffic can go inside and what can leave the system or network. IPS blocks any network traffic that matches present rules and attack signatures.
- Setting up logging and monitoring devices: Proper network logging and monitoring are essential for detecting malicious activities and intrusions. If a new unauthorized device appears on our network, we should be able to detect it.

A Security Operations Center (SOC) is a team of cyber security professionals that monitors the network and its systems to detect malicious cyber security events

Digital Forensics and Incident Response (DFIR) covers:

- Digital Forensics
- Incident Response
- Malware Analysis

Malware stands for malicious software. Software refers to programs, documents, and files you can save on a disk or send over the network.

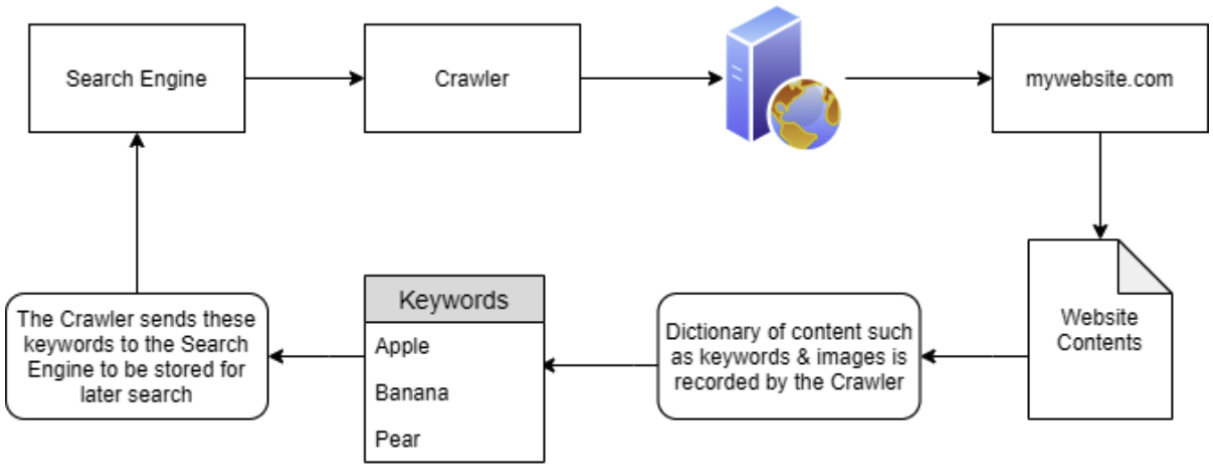
Malware analysis aims to learn about such malicious programs using various means:

- Static analysis works by inspecting the malicious program without running it. This usually requires solid knowledge of assembly language (the processor's instruction set, i.e., the computer's fundamental instructions).

- Dynamic analysis works by running the malware in a controlled environment and monitoring its activities. It lets you observe how the malware behaves when running.

Google Dorking

"Search Engines" such as Google are huge indexers – specifically, indexers of content spread across the World Wide Web. These essentials in surfing the internet use “Crawlers” or “Spiders” to search for this content across the World Wide Web. These crawlers discover content through various means. One being by pure discovery, where a URL is visited by the crawler and information regarding the content type of the website is returned to the search engine. The diagram below is a high-level abstraction of how these web crawlers work. Once a web crawler discovers a domain such as mywebsite.com, it will index the entire contents of the domain, looking for keywords and other miscellaneous information. Crawlers attempt to traverse, termed as crawling, every URL and file that they can find



Search Engine Optimisation or SEO is a prevalent and lucrative topic in modern-day search engines. There are many factors in how “optimal” a domain is - resulting in something like a point-scoring system. To highlight a few influences on how these points are scored, factors such as:

- • How responsive your website is to the different browser types I.e. Google Chrome, Firefox and Internet Explorer - this includes Mobile phones!
- • How easy it is to crawl your website using "Sitemaps"

Robots.txt: this file is the first thing indexed by "Crawlers" when visiting a website. This file must be served at the root directory - specified by the webserver itself. A very basic markup of a Robots.txt is like the following:

```
User-agent: *
Allow: /

Sitemap: http://mywebsite.com/sitemap.xml

```

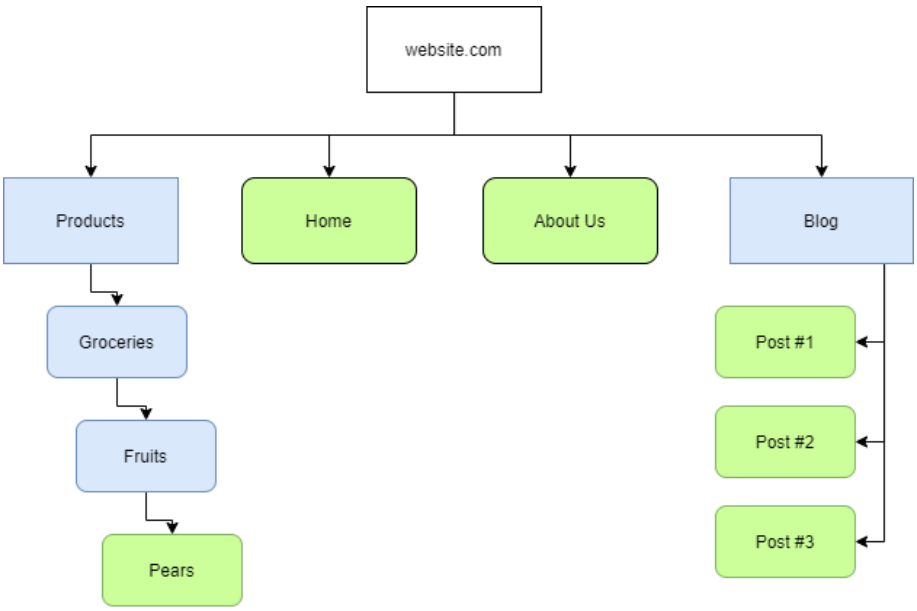
Here we have a few keywords...

Keyword	Function
User-agent	Specify the type of "Crawler" that can index your site (the asterisk being a wildcard, allowing **all "User-agents"**
Allow	Specify the directories or file(s) that the "Crawler" **can** index
Disallow	Specify the directories or file(s) that the "Crawler" **cannot** index
Sitemap	Provide a reference to where the sitemap is located (improves SEO as previously discussed, we'll come to sitemaps in the next task)

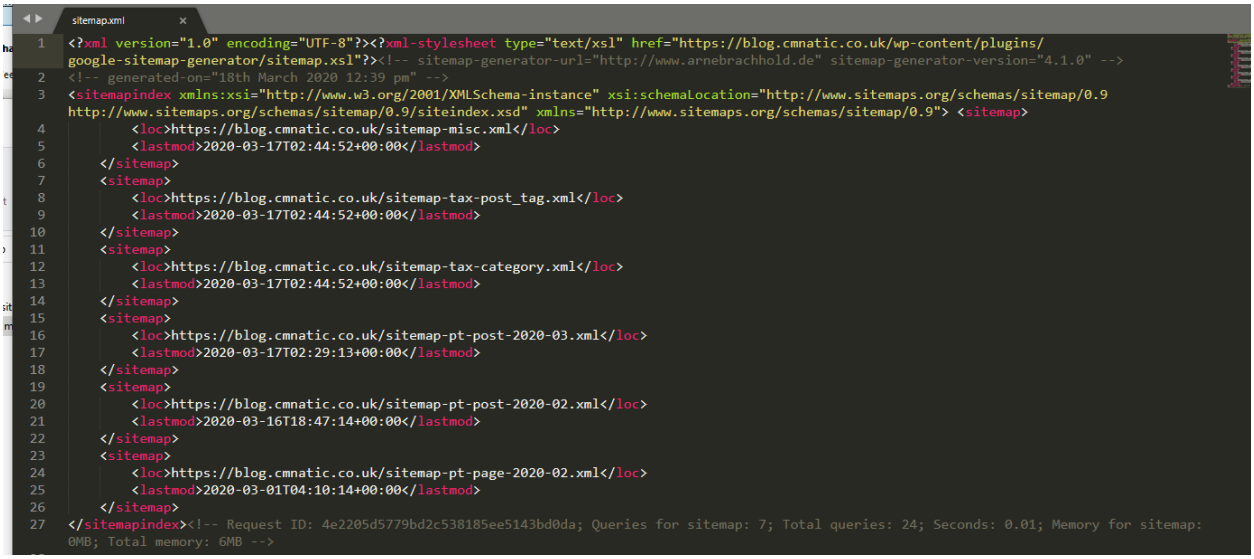
Robots.txt works on a "blacklisting" basis. Essentially, unless told otherwise, the Crawler will index whatever it can find.

Sitemaps:

“Sitemaps” are indicative resources that are helpful for crawlers, as they specify the necessary routes to find content on the domain.



The blue rectangles represent the route to nested-content, similar to a directory I.e. “Products” for a store. Whereas, the green rounded-rectangles represent an actual page. However, this is for illustration purposes only - “Sitemaps” don't look like this in the real world.



“Sitemaps” are XML formatted. The presence of "Sitemaps" holds a fair amount of weight in influencing the "optimisation" and favorability of a website. Search engines have a lot of data to process. The efficiency of how this data is collected is paramount. Resources like "Sitemaps" are extremely helpful for "Crawlers" as the necessary routes to content are already provided.

Using Google for Advanced Searching

We can use terms such as “site” (such as bbc.co.uk) and a query (such as "gchq news") to search the specified site for the keyword we have provided to filter out content that may be harder to find otherwise

A few common terms we can search and combine include:

Term	Action
filetype:	Search for a file by its extension (e.g. PDF)
cache:	View Google's Cached version of a specified URL
intitle:	The specified phrase MUST appear in the title of the page